Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Virginia Estellers
Exercises: Emanuel Laude
Winter Semester 2017/18

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 8

Room: 02.09.023
Friday, 19.01.2018, 09:15-11:00
Submission deadline: Monday, 15.01.2018, 10:15, Room 02.09.023

Theory: Stochastic Gradient Descent (8 Points)

Exercise 1 (8 Points)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be L-smooth, i.e. differentiable with Lipschitz-continuous gradient with modulus $L > 0$. Let $x, y \in \mathbb{R}^n$. Prove that it holds

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|x - y\|_2^2.$$

Hint: First prove that it holds that:

$$f(y) = f(x) + \int_0^1 \frac{\partial f(x + t(y - x))}{\partial t} \, \mathrm{d}t.$$

Coding: Multinomial Logistic Regression on MNIST with SGD (16 Points)

In this exercise you are asked to implement different variants of (minibatch-)SGD for logistic regression, as discussed in the lecture. Benchmark the different variants against each other on MNIST. You may use the solution for the logistic regression task (sheet 4) with full batch gradient descent, provided online. More precisely you are asked to do the following:

- Download the solution for full batch proximal gradient descent for logistic regression and alter the code accordingly.
- Play around with different batch sizes including batch size one (original SGD).
- Play around with different step sizes (learning rates) σ_t, including a vanishing step size schedule $\sigma_t := \frac{C}{t}$, for some $C > 0$.
- Benchmark the different variants (including full batch and original SGD with vanishing step size) against each other on MNIST, plot the energy over the iterations.